

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-228487

(43) 公開日 平成10年(1998) 8月25日

(51) Int.Cl.\*

識別記号

F I

G 0 6 F 17/30

G 0 6 F 15/403

3 4 0 A

15/401

3 3 0 Z

審査請求 有 請求項の数 5 F D (全 16 頁)

(21) 出願番号 特願平9-48512

(22) 出願日 平成9年(1997) 2月17日

(71) 出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 安倍 直樹

東京都港区芝五丁目7番1号 日本電気株

式会社内

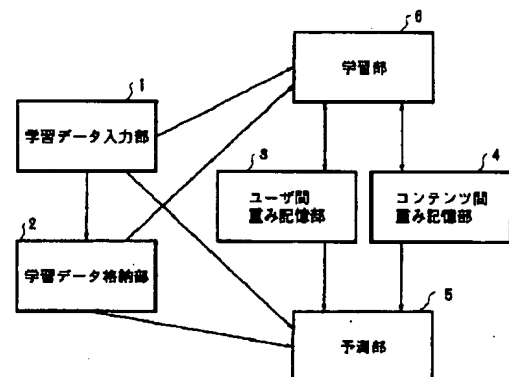
(74) 代理人 弁理士 境 廣巳

(54) 【発明の名称】 ユーザ嗜好自動学習方式

(57) 【要約】

【課題】 コンテンツに対するユーザの嗜好を、データとして与えられたユーザ・コンテンツ対とその正しい嗜好度を用いて効率良く学習する。

【解決手段】 ユーザ・コンテンツ対が入力されると、予測部5は、その対の嗜好度を、現在の仮説(学習データ格納部2に格納された過去の学習データの嗜好度の値、記憶部3、4に格納されているユーザ間重み、コンテンツ間重み)によって予測する。学習部6は、その対の正しい嗜好度が入力されると、現在の仮説を更新する。その際、ユーザ間重みは、過去の学習データ中のユーザ・コンテンツ対に対する嗜好度と新たに与えられたユーザ・コンテンツ対に対する嗜好度との差と、前記コンテンツ間重みとの関数として更新し、コンテンツ間重みは、過去の学習データ中のユーザ・コンテンツ対に対する嗜好度と新たに与えられたユーザ・コンテンツ対に対する嗜好度との差と、前記ユーザ間重みとの関数として更新する。



## 【特許請求の範囲】

【請求項1】 ニュース記事やホームページ等のコンテンツに対するユーザの嗜好を、データとして与えられたコンテンツとそのコンテンツに対するユーザの嗜好度を表す数値からなるデータを用いて学習する装置において、ユーザ・コンテンツ対および該コンテンツに対する該ユーザの嗜好度を表す数値からなる学習データを入力する学習データ入力部と、過去に与えられた学習データを一括して格納する学習データ格納部と、ユーザ間の類似度を表すユーザ間重みを格納するユーザ間重み記憶部と、コンテンツ間の類似度を表すコンテンツ間重みを格納するコンテンツ間重み記憶部と、前記学習データ入力部に与えられたユーザ・コンテンツ対に対して、前記学習データ格納部に格納された学習データ、前記ユーザ間重み記憶部に格納されたユーザ間重み、および前記コンテンツ間重み記憶部に格納されたコンテンツ間重みを用いて、そのユーザのそのコンテンツに対する嗜好度を予測する予測部と、前記学習データ入力部から入力された前記ユーザ・コンテンツ対の正しい嗜好度、前記学習データ格納部に格納された過去に与えられた学習データ、前記ユーザ間重み記憶部に格納されたユーザ間重み、および前記コンテンツ間重み記憶部に格納されたコンテンツ間重みを用いて、前記ユーザ間重み記憶部に格納されたユーザ間重みおよび前記コンテンツ間重み記憶部に格納されたコンテンツ間重みを更新する学習部とを備え、前記予測部における予測は、前記学習データ格納部に格納される過去の学習データの嗜好度の値の、新たに与えられたユーザ・コンテンツ対との間のユーザ間重みとコンテンツ間重みの積を重みとする、重みつき平均によって行われる構成を有し、前記学習部における重み更新は、過去の学習データ中の各ユーザ・コンテンツ対に対して、新たに与えられたユーザ・コンテンツ対との間のユーザ間重みとコンテンツ間重みとの双方に対して行われ、且つ、ユーザ間重みは、過去の学習データ中のユーザ・コンテンツ対に対する嗜好度と新たに与えられたユーザ・コンテンツ対に対する嗜好度との差と、前記コンテンツ間重みとの関数として更新され、コンテンツ間重みは、過去の学習データ中のユーザ・コンテンツ対に対する嗜好度と新たに与えられたユーザ・コンテンツ対に対する嗜好度との差と、前記ユーザ間重みとの関数として更新される構成を有することを特徴とするユーザ嗜好自動学習方式。

【請求項2】 請求項1記載の学習部において、過去の学習データ中の各ユーザ・コンテンツ対に対して、新たに与えられたユーザ・コンテンツ対との間のユーザ間重みとコンテンツ間重みとの双方を更新するのに代えて、

乱数を用いてランダムに、ユーザ間重み又はコンテンツ間重みを更新する構成を有することを特徴とする請求項1記載のユーザ嗜好自動学習方式。

【請求項3】 ニュース記事やホームページ等のコンテンツに対するユーザの嗜好を、データとして与えられたコンテンツとそのコンテンツに対するユーザの嗜好度を表す数値からなるデータを用いて学習する装置において、

ユーザ・コンテンツ対および該コンテンツに対する該ユーザの嗜好度を表す数値からなる学習データを入力する学習データ入力部と、

過去に与えられた学習データを一括して格納する学習データ格納部と、

複数の学習方式を格納する学習方式記憶部と、

前記学習データ入力部に与えられたユーザ・コンテンツ対に対して、そのユーザのそのコンテンツに対する嗜好度を、前記学習方式記憶部に格納された各学習方式により予測させ、それらの予測値の、各学習方式に対する信頼度の重み付き平均値をもって予測する予測部と、

複数のユーザ・コンテンツ対からなる質問対候補集合の各ユーザ・コンテンツ対について、そのユーザのそのコンテンツに対する嗜好度を前記学習方式記憶部中の各学習方式により予測させて、最も予測値のばらつきの大きいユーザ・コンテンツ対の正しい嗜好度を質問してその値を得る質問部と、

前記学習データ入力部から入力された前記ユーザ・コンテンツ対の正しい嗜好度および前記質問部が質問して得た前記ユーザ・コンテンツ対の正しい嗜好度に基づいて、

前記学習方式記憶部中の各学習方式の保持する仮説を更新すると共に、前記予測部が使用する各学習方式に対する信頼度を表す重みを、各重みの現在値および各学習方式の予測値と入力として与えられた正しい嗜好度との差の関数として更新する学習部とを有することを特徴とするユーザ嗜好自動学習方式。

【請求項4】 前記学習方式記憶部に記憶される各学習方式が、請求項2記載のユーザ嗜好自動学習方式である請求項3記載のユーザ嗜好自動学習方式。

【請求項5】 前記質問部は、前記複数のユーザ・コンテンツ対からなる質問対候補集合として、過去の学習データ中に現れるユーザとコンテンツとからなるユーザ・コンテンツ対の中で、過去にデータとして嗜好度を与えられていない対の集合の全て又は一部を用いる構成を有することを特徴とする請求項4記載のユーザ嗜好自動学習方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、コンピュータネットワーク上で、ニュース記事やホームページ等のコンテンツ・サービスにおいて、個々のユーザの嗜好に合ったコンテンツを提供するためのユーザ嗜好の学習技術に

関するものである。

【0002】

【従来の技術】コンピュータネットワーク上で、ニュース記事やホームページ等のコンテンツに対するユーザの嗜好を学習し、個々のユーザの嗜好に合ったコンテンツを提供する学習型ナビゲーション・システムにおいて、コンピュータネットワーク上のコンテンツに対する嗜好の学習法として従来知られている方法は、同一ユーザによる類似のコンテンツに対する過去の嗜好を基に予測・学習する「コンテンツベース方式」と、同一コンテンツに対する類似ユーザの嗜好を基に予測・学習する「ユーザベース方式」とに大別できる。

【0003】前者のコンテンツベース方式の例としては、例えばカーネギーメロン大学のKen Langによるニュース・ウィーダー (NewsWeeder) が知られており、Proceedings of the 12th International Conference on Machine Learningに掲載の論文「NewsWeeder: Learning to filter netnews」に詳細に掲載されている。また、同じくコンテンツベース方式の例として、日本電気株式会社C&C研究所の中村篤祥等による逐次型学習方式もあり、第52回情報処理学会全国大会の予稿集掲載の論文「ブール変数実数多項式を用いた嗜好関数の学習」に詳細に記載されている。

【0004】後者のユーザベース方式の例としては、MITのPaul Resnick等による「グループ・レンズ (Group Lens)」が知られており、Proceedings of CSCW (1994) 掲載の論文「GroupLens: An Open Architecture for Collaborative Filtering of Netnews」に詳細に記載されている。

【0005】

【発明が解決しようとする課題】近年、上記の二種類の方式を融合し、類似するユーザの過去の嗜好の情報と、類似するコンテンツに対する過去の嗜好の情報の両方を利用し、ユーザの嗜好を学習・予測する方式の重要性が強く認識されており、これを達成する有効な方式の発明が待望されていた。本発明は、まさにこの条件を満たす、具体的かつ有効な方式を提案するものである。

【0006】

【課題を解決するための手段】

(A) 本発明のユーザ嗜好自動学習方式は、ニュース記事やホームページ等のコンテンツに対するユーザの嗜好を、データとして与えられたコンテンツとそのコンテンツに対するユーザの嗜好度を表す数値からなるデータを用いて学習する装置において、ユーザ・コンテンツ対および該コンテンツに対する該ユーザの嗜好度を表す数値からなる学習データを入力する学習データ入力部と、過

去に与えられた学習データを一括して格納する学習データ格納部と、ユーザ間の類似度を表すユーザ間重みを格納するユーザ間重み記憶部と、コンテンツ間の類似度を表すコンテンツ間重みを格納するコンテンツ間重み記憶部と、前記学習データ入力部に与えられたユーザ・コンテンツ対に対して、前記学習データ格納部に格納された学習データ、前記ユーザ間重み記憶部に格納されたユーザ間重み、および前記コンテンツ間重み記憶部に格納されたコンテンツ間重みを用いて、そのユーザのそのコンテンツに対する嗜好度を予測する予測部と、前記学習データ入力部から入力された前記ユーザ・コンテンツ対の正しい嗜好度、前記学習データ格納部に格納された過去に与えられた学習データ、前記ユーザ間重み記憶部に格納されたユーザ間重み、および前記コンテンツ間重み記憶部に格納されたコンテンツ間重みを用いて、前記ユーザ間重み記憶部に格納されたユーザ間重みおよび前記コンテンツ間重み記憶部に格納されたコンテンツ間重みを更新する学習部とを備え、前記予測部における予測は、前記学習データ格納部に格納される過去の学習データの嗜好度の値の、新たに与えられたユーザ・コンテンツ対との間のユーザ間重みとコンテンツ間重みの積を重みとする、重みつき平均によって行われる構成を有し、前記学習部における重み更新は、過去の学習データ中の各ユーザ・コンテンツ対に対して、新たに与えられたユーザ・コンテンツ対との間のユーザ間重みとコンテンツ間重みとの双方に対して行われ、且つ、ユーザ間重みは、過去の学習データ中のユーザ・コンテンツ対に対する嗜好度と新たに与えられたユーザ・コンテンツ対に対する嗜好度との差と、前記コンテンツ間重みとの関数として更新され、コンテンツ間重みは、過去の学習データ中のユーザ・コンテンツ対に対する嗜好度と新たに与えられたユーザ・コンテンツ対に対する嗜好度との差と、前記ユーザ間重みとの関数として更新される構成を有する。

【0007】上記(A)のように構成された本発明のユーザ嗜好自動学習方式にあっては、学習データ入力部からユーザ・コンテンツ対が入力されると、予測部が、そのユーザ・コンテンツ対に対して、学習データ格納部に格納された過去の学習データ、ユーザ間重み記憶部に格納されたユーザ間重み、およびコンテンツ間重み記憶部に格納されたコンテンツ間重みを現在の仮説として、そのユーザのそのコンテンツに対する嗜好度の予測値として、過去の学習データの嗜好度の値の、新たに与えられたユーザ・コンテンツ対との間のユーザ間重みとコンテンツ間重みの積を重みとする、重みつき平均によって求める。また、そのユーザ・コンテンツ対の正しい嗜好度が学習データ入力部から入力されると、学習部が、この正しい嗜好度、学習データ格納部に格納された過去に与えられた学習データ、ユーザ間重み記憶部に格納された現在のユーザ間重み、およびコンテンツ間重み記憶部に格納された現在のコンテンツ間重みを用いて、ユーザ間

重み記憶部に格納されたユーザ間重みおよびコンテンツ間重み記憶部に格納されたコンテンツ間重み、つまり現在の仮説を更新する。具体的には、過去の学習データ中の各ユーザ・コンテンツ対に対して、新たに与えられたユーザ・コンテンツ対との間のユーザ間重みとコンテンツ間重みとの双方を更新し、その際、ユーザ間重みは、過去の学習データ中のユーザ・コンテンツ対に対する嗜好度と新たに与えられたユーザ・コンテンツ対に対する嗜好度との差と、前記コンテンツ間重みとの関数として更新し、コンテンツ間重みは、過去の学習データ中のユーザ・コンテンツ対に対する嗜好度と新たに与えられたユーザ・コンテンツ対に対する嗜好度との差と、前記ユーザ間重みとの関数として更新する。このような重みの更新を行うことにより、ユーザ間の類似度のみを考慮する場合や、コンテンツ間の類似度のみを考慮する場合に比べて、立上りの早い学習が可能となる。なお、以上の動作例では、ユーザ・コンテンツ対が与えられたときに、その予測と学習とを行うようにしたが、これは毎回学習を行ういわゆる逐次学習式の動作であり、別の動作例として、学習中は予測を行わずに学習だけを行い、或る程度学習が進んだ段階以降は、予測だけを行う所謂バッチ式の動作を行わせることもできる。

【0008】(B) また本発明の別のユーザ嗜好自動学習方式は、前記の学習部において、過去の学習データ中の各ユーザ・コンテンツ対に対して、新たに与えられたユーザ・コンテンツ対との間のユーザ間重みとコンテンツ間重みとの双方を更新するのに代えて、乱数を用いてランダムに、ユーザ間重み又はコンテンツ間重みを更新する構成を有する。

【0009】(C) 更に本発明の別のユーザ嗜好自動学習方式は、ニュース記事やホームページ等のコンテンツに対するユーザの嗜好を、データとして与えられたコンテンツとそのコンテンツに対するユーザの嗜好度を表す数値からなるデータを用いて学習する装置において、ユーザ・コンテンツ対および該コンテンツに対する該ユーザの嗜好度を表す数値からなる学習データを入力する学習データ入力部と、過去に与えられた学習データを一括して格納する学習データ格納部と、複数の学習方式を格納する学習方式記憶部と、前記学習データ入力部に与えられたユーザ・コンテンツ対に対して、そのユーザのそのコンテンツに対する嗜好度を、前記学習方式記憶部に格納された各学習方式により予測させ、それらの予測値の、各学習方式に対する信頼度を表す重み付き平均値をもって予測する予測部と、複数のユーザ・コンテンツ対からなる質問候補集合の各ユーザ・コンテンツ対について、そのユーザのそのコンテンツに対する嗜好度を前記学習方式記憶部中の各学習方式により予測させて、最も予測値のばらつきの大きいユーザ・コンテンツ対の正しい嗜好度を質問してその値を得る質問部と、前記学習データ入力部から入力された前記ユーザ・コンテンツ対

の正しい嗜好度および前記質問部が質問して得た前記ユーザ・コンテンツ対の正しい嗜好度に基づいて、前記学習方式記憶部中の各学習方式の保持する仮説を更新すると共に、前記予測部が使用する各学習方式に対する信頼度を表す重みを、各重みの現在値および各学習方式の予測値と入力として与えられた正しい嗜好度との差の関数として更新する学習部とを有する。

【0010】上記(C)のように構成された本発明のユーザ嗜好自動学習方式にあっては、学習データ入力部からユーザ・コンテンツ対が入力されると、予測部が、そのユーザ・コンテンツ対に対して、そのユーザのそのコンテンツに対する嗜好度を、学習方式記憶部に格納された各学習方式により予測させてその予測値の重み付き平均値をもって予測し、そのユーザ・コンテンツ対の正しい嗜好度が学習データ入力部から入力されると、学習部が、その正しい嗜好度に基づいて、学習方式記憶部中の各学習方式の保持する仮説を更新すると共に、予測部が使用する各学習方式に対する信頼度を表す重みを、各重みの現在値および各学習方式の予測値と入力として与えられた正しい嗜好度との差の関数として更新する。このように、各学習方式の仮説の更新と共に、各学習方式に対する信頼度を表す重みの更新を行うことにより、学習が進むにつれ、より信頼度の高い学習方式の予測値が大きい影響力を持つようになり、少ない学習データによって高い精度の予測が可能となる。

【0011】他方、質問部は、システム自らが学習すべきユーザ・コンテンツ対を決定することにより、学習すべきユーザ・コンテンツ対を入力する利用者の手間を削減すると同時に、より効率的な学習を可能にする。つまり、質問部は、複数のユーザ・コンテンツ対からなる質問候補集合中の各ユーザ・コンテンツ対について、そのユーザのそのコンテンツに対する嗜好度を学習方式記憶部中の各学習方式により予測させて、最も予測値のばらつきの大きいユーザ・コンテンツ対の正しい嗜好度を質問してその値を取得する。そして、学習部が、質問部が質問したユーザ・コンテンツ対について利用者が回答した正しい嗜好度に基づいて、学習方式記憶部中の各学習方式の保持する仮説を更新すると共に、予測部が使用する各学習方式に対する信頼度を表す重みを、更新する。このように情報量の高いユーザ・コンテンツ対に基づいて学習を進めることにより、少ない質問数で高い精度の予測が可能となる。

【0012】学習の順序は、学習データ入力部から1つのユーザ・コンテンツ対を入力してその予測を行った後に、正しい嗜好度を学習データ入力部から入力して学習するといった逐次式であっても良く、学習中は予測を行わずに学習だけを行い、或る程度学習が進んだ段階以降は、予測だけを行うバッチ式であっても良い。また、学習データ入力部からの入力と質問部による質問との関係については、学習データ入力部から1つのユーザ・コン

テンツ対の入力を受け、それを処理した後に、質問部で可能であれば1つの質問を出すようにしても良く、複数のユーザ・コンテンツ対の入力を学習データ入力部から受け付け、それを処理した後に、質問部から幾つかの質問を出すようにしても良い。

【0013】(D)更に本発明の別のユーザ嗜好自動学習方式は、上記(C)におけるユーザ嗜好自動学習方式において、学習方式記憶部に格納する学習方式に上記(B)で述べたユーザ嗜好自動学習方式を使用する。つまり、学習方式記憶部に格納される個々の学習方式は、ユーザ間の類似度を表すユーザ間重みを格納するユーザ間重み記憶部と、コンテンツ間の類似度を表すコンテンツ間重みを格納するコンテンツ間重み記憶部と、予測部と、学習部とを含む。個々の学習方式における予測部は、与えられたユーザ・コンテンツ対に対して、学習データ格納部に格納された学習データ、ユーザ間重み記憶部に格納されたユーザ間重み、およびコンテンツ間重み記憶部に格納されたコンテンツ間重みを用いて、そのユーザのそのコンテンツに対する嗜好度を予測する。その際、予測は、学習データ格納部に格納される過去の学習データの嗜好度の値の、新たに与えられたユーザ・コンテンツ対との間のユーザ間重みとコンテンツ間重みの積を重みとする、重みつき平均によって行われる。また、個々の学習方式における学習部は、ユーザ・コンテンツ対の正しい嗜好度、学習データ格納部に格納された過去に与えられた学習データ、ユーザ間重み記憶部に格納されたユーザ間重み、およびコンテンツ間重み記憶部に格納されたコンテンツ間重みを用いて、ユーザ間重み記憶部に格納されたユーザ間重み、コンテンツ間重み記憶部に格納されたコンテンツ間重みを更新する。その際、乱数を用いてランダムに、ユーザ間重み又はコンテンツ間重みを更新する。すなわち、学習方式記憶部中の学習方式はそれぞれ同一の学習方式であるが、一般に、ユーザ間重み、コンテンツ間重みの更新にあたって異なる乱数が発生され、異なる重みの更新が行われていくので、実質的には、特定の場面において異なる予測を行う複数の学習方式となる。

【0014】また更に本発明の別のユーザ嗜好自動学習方式は、上記(D)のユーザ嗜好自動学習方式において、前記質問部は、前記複数のユーザ・コンテンツ対からなる質問候補集合として、過去の学習データ中に現れるユーザとコンテンツとからなるユーザ・コンテンツ対の中で、過去にデータとして嗜好度を与えられていない対の集合の全て又は一部を用いる構成を有する。

【0015】

【発明の実施の形態】次に本発明の実施の形態の例について図面を参照して詳細に説明する。

【0016】図1を参照すると、本発明のユーザ嗜好自動学習方式の第1の実施の形態は、学習データ入力部1と、学習データ格納部2と、ユーザ間重み記憶部3と、

コンテンツ間重み記憶部4と、予測部5と、学習部6とから構成されている。これらは、例えばプログラム制御されたCPU、キーボード等の入力装置、磁気ディスク等の記憶装置を有するデータ処理装置(コンピュータ)で実現することができる。

【0017】学習データ入力部1は、学習データを逐次的に入力する部分である。入力される学習データは、ユーザ・コンテンツ対と、そのコンテンツに対するそのユーザの正しい嗜好度を表す数値とから構成される。

【0018】学習データ格納部2は過去に与えられた学習データを一括して格納する部分、ユーザ間重み記憶部3はユーザ間の類似度を表す重み行列(ユーザ間重みの行列)を格納する部分、コンテンツ間重み記憶部4はコンテンツ間の類似度を表す重み行列(コンテンツ間重みの行列)を格納する部分である。

【0019】予測部5は、学習データ入力部1から入力されたユーザ・コンテンツ対について、そのユーザのそのコンテンツに対する嗜好度を予測する部分である。この予測は、学習データ格納部2に格納されている過去の学習データと、ユーザ間重み記憶部3に格納されているユーザ間重みと、コンテンツ間重み記憶部4に格納されているコンテンツ間重みを参照して行われる。

【0020】学習部6は、学習データ入力部1から新たな学習データ(ユーザ・コンテンツ対とその正しい嗜好度を表す数値)が入力された場合に、入力されたユーザ・コンテンツ対の正しい嗜好度、学習データ格納部2に格納された過去に与えられた学習データ、ユーザ間重み記憶部3に格納されたユーザ間重み、およびコンテンツ間重み記憶部4に格納されたコンテンツ間重みを用いて、ユーザ間重み記憶部3に格納されたユーザ間重み及びコンテンツ間重み記憶部4に格納されたコンテンツ間重みを更新する部分である。

【0021】図2は本実施の形態の動作の一例を示すフローチャートである。この動作例は、学習データ入力部1から1つのユーザ・コンテンツ対を入力し、そのコンテンツに対するそのユーザの嗜好度を予測部5において予測し、学習データ入力部1からそのユーザ・コンテンツ対の正しい嗜好度を入力し、学習部6においてユーザ間重み及びコンテンツ間重みを更新する、という処理を繰り返す逐次式の学習を採った場合のものである。

【0022】いま、ユーザの集合をX、コンテンツの集合をYとすると、本発明の第1の実施の形態においては、以下のようなステップを実行する。

(1)新しいユーザ $x \in X$ と新しいコンテンツ $y \in Y$ の対を学習データ入力部1にて入力する(ステップS1)。

(2)入力されたユーザ $x$ とコンテンツ $y$ の対について、ユーザ $x$ のコンテンツ $y$ に対する嗜好度 $f(x, y)$ を、現在の仮説を用いて予測部5で予測する(ステップS2)。 $f(x, y)$ の値は一般に実数値であると

仮定する。例えば、5段階評価(1, 2, ..., 5)等が想定される。

(3) ユーザ $x$ のコンテンツ $y$ に対する正しい嗜好度 $f(x, y)$ の値を学習データ入力部1にて入力し(ステップS3)、学習データ格納部2に格納する(ステップS4)。

(4) 学習部6において現在の仮説を更新(学習)する(ステップS5)。

(5) 学習を終了するか否かを判定し、終了しない場合はステップS1に戻る(ステップS6)。

【0023】予測・学習アルゴリズムの目標は、なるべく少ない数の学習データで、高い精度の予測を行うことにある。ステップS6で終了とする条件(停止条件)としては様々な条件が考えられるが、本発明については本質的な影響はなく、また一般に適用環境に大きく依存するので、ここでは特に特定しない。なお、上記(1)~(4)を無限なく繰り返すようにしても良い。

【0024】上記(2)の予測段階では、以下のような仮説形式を用いる。まず、過去に与えられた正しい嗜好度 $f(x, y)$ の値がデータ行列 $M$ として学習データ格納部2に格納されているものとする。このとき、過去に $f(x, y)$ の値が与えられた $x, y$ については、 $M$  \*  

$$\hat{f}(x, y) = \frac{\sum_{M(x', y') \neq *}{u(x, x') \cdot v(y, y') \cdot M(x', y')}{\sum_{M(x', y') \neq *} u(x, x') \cdot v(y, y')} \dots \dots \dots (1)$$

【0026】重みとしてユーザ間の重みとコンテンツ間の重みの積を用いているのは、以下のような意味がある。 $f(x', y')$ の値が $f(x, y)$ の値と近い値をとるのは、 $x, x'$ と、 $y, y'$ が両方とも高い類似度を持つ場合に限るので、 $(x, y)$ と $(x', y')$ が類似するという命題は、 $x, x'$ が類似するという命題と、 $y, y'$ が類似するという命題との論理積であると考えられる。しかし、ノイズを含む環境からのしかも不完全なデータから学習を行う場合には論理的な推論は危険なので、ここでは、論理積の代わりに、類似度を表す重みによる実数積を用いている。そうすることにより、ノイズに対する頑健性を備えた予測を実現している。

【0027】次に学習の手順、すなわち現在の仮説を更新する手順について説明する。今、新しい対 $x, y$ に対する嗜好度の正解 $f(x, y)$ を与えられたとする。過去に与えられた各正解 $f(x', y')$ に対して、もし $|f(x, y) - f(x', y')|$ が大きい値をとるならば、 $(x, y)$ と $(x', y')$ が類似しているという命題が否定されることになる。上記の観察から、これは、 $(x, x')$ が類似しているという命題と $(y, y')$ が類似しているという命題の論理積が否定されたことになるから、 $(x, x')$ が類似していないという

\* $(x, y) = f(x, y)$ であり、そうでない $x, y$ については、 $M(x, y) = *$ である。ここで、\*は未設定であることを示す。次に、任意のユーザの対 $(x_1, x_2) \in X \times X$ に対して、それらユーザ間の類似度を表す重み $u(x_1, x_2)$ を記述したユーザ間の重み行列 $u$ を、ユーザ間重み記憶部3に設ける。同様に、任意のコンテンツの対 $(y_1, y_2) \in Y \times Y$ に対して、それらコンテンツ間の類似度を表す重み $v(y_1, y_2)$ を記述したコンテンツ間の重み行列 $v$ を、コンテンツ間重み記憶部4に設ける。現在の仮説とは、上記のデータ行列 $M$ と、ユーザ間の重み行列 $u$ と、コンテンツ間の重み行列 $v$ からなる。そして、新たに与えられた $(x, y) \in X \times Y$ に対する嗜好度の予測は、データ行列 $M$ に格納された全データによる重み付き平均をもって行う。ここで、重み付き平均をとる場合の重みとしては、新しいユーザ、コンテンツ対を $x, y$ 、データ行列 $M$ 中のユーザ、コンテンツ対を $x', y'$ とすると、ユーザ間の重みとコンテンツ間の重みの積、 $u(x, x') \cdot v(y, y')$ を用いる。すなわち、新しい対 $(x, y)$ の嗜好度の予測値 $\hat{f}(x, y)$ を以下のように計算する。

【0025】

【数1】

命題と $(y, y')$ が類似していないという命題のどちらかが正しいことになる。しかし、そのどちらが正しいかは、これだけの情報からは特定できない。そこで、過去のデータにより学習された $(x, x')$ の類似度の推定値 $u(x, x')$ と、 $(y, y')$ の類似度の推定値 $v(y, y')$ を参照して以下のような推論を行う。すなわち、もし $u(x, x')$ の値が大きく $(x, x')$ が類似しているという証拠が強いならば、新しく与えられた命題は、 $(y, y')$ が類似していないという命題を指示し、また逆に、 $v(y, y')$ の値が大きく $(y, y')$ が類似しているという証拠が強いならば、 $(x, x')$ が類似していないという命題を指示する。

上記の「大きい」や「強い」という判断は、閾値等を用いて論理値として推論することも可能であるが、ノイズを含んだ環境からの不完全なデータを用いた学習においてはそれは危険であるので、 $|f(x, y) - f(x', y')|$ の大きさと $u(x, x')$ の値の大きさによって連続的に $v(y, y')$ の更新幅を決定し、また $|f(x, y) - f(x', y')|$ の大きさと $v(y, y')$ の値の大きさによって連続的に $u(x, x')$ の更新幅を決定する。例えば、ユーザ間の重み $u$ は以下のように更新し、

【0028】

【数2】

$$u(x, x') := u(x, x') \cdot e^{-c \cdot v(y, y') \cdot (|f(x', y') - f(x, y)| - d)} \quad \dots \dots (2)$$

【0029】コンテンツ間の重み  $v$  は以下のように更新 \* 【0030】

する。

\* 【数3】

$$v(y, y') := v(y, y') \cdot e^{-c \cdot u(x, x') \cdot (|f(x', y') - f(x, y)| - d)} \quad \dots \dots (3)$$

【0031】なお、上式の  $c$ 、 $d$  は正の定数である。  $c$  ※ 変化量の上限を指定することも可能である。例えば、ユーザ間の重み  $u$  は以下のように更新し、

とも可能である。  $d$  には  $|f(x', y') - f(x, y)|$  の平均値の推定値等を用いるのが適当である。ま 10 【0032】

た、以上の更新を行う際に、一回の更新における重みの※ 【数4】

$$u(x, x') := u(x, x') \cdot \max\{0.5, \min\{2.0, e^{-c \cdot v(y, y') \cdot (|f(x', y') - f(x, y)| - d)}\}\} \quad \dots \dots (4)$$

【0033】コンテンツ間の重み  $v$  は以下のように更新 ★ 【0034】

しても良い。

★ 【数5】

$$v(y, y') := v(y, y') \cdot \max\{0.5, \min\{2.0, v(y, y') \cdot e^{-c \cdot u(x, x') \cdot (|f(x', y') - f(x, y)| - d)}\}\} \quad \dots \dots (5)$$

【0035】また、アンダーフロー等の問題を避けるた ☆ 【0036】

めに、重みの更新を行う度に、以下のように正規化を行 【数6】

っても良い。 ☆

$$u(x, x') := \frac{u(x, x')}{\sum_{x'} u(x, x')} \quad \dots \dots (6)$$

【0037】

◆ ◆ 【数7】

$$v(y, y') := \frac{v(y, y')}{\sum_{y'} v(y, y')} \quad \dots \dots (7)$$

【0038】上のような重みの更新法を用いることにより、ノイズに対する頑健な帰納推論を実現することが可能になると同時に、ユーザ間の類似度のみを考慮する場合や、コンテンツ間の類似度のみを考慮する場合に比べて、立上がりの早い学習が可能となる。

【0039】次に、嗜好度の予測および重み行列の更新の具体例を説明する。例としては、図4に示すように  $u_1$  から  $u_5$  までの5人のユーザと  $c_1$  から  $c_7$  までの7つのコンテンツとを想定し、学習データ格納部2における過去の学習データを示すデータ行列  $M$  の要素が全て未設定（図4では？で示す）である状態において、 $f(u_2, c_3) = 4$ 、 $f(u_4, c_3) = 4$ 、 $f(u_2, c_6) = 1$ 、 $f(u_4, c_6) = 1$  の順序で学習データが与えられたものとする。また、ユーザ間重み記憶部3に設けられているユーザ間重み及びコンテンツ間重み記憶

部4に設けられているユーザ間重みはそれぞれ初期値1に設定されているものとする。更に、定数  $c$  は、 $-\log 0.8$  に設定、すなわち  $e^{-c} = 0.8$  であり、 $d$  は2であるとする。また、自分自身の重み  $u(x, x)$ 、 $v(y, y)$  は更新しないものとする。このとき、各学習データが与えられた後の嗜好度の予測値および重みの更新は以下ようになる。

【0040】1.  $f(u_2, c_3) = 4$

この場合、過去の学習データが存在しないため、重みの更新はない。また、嗜好度の予測値は任意のものとなる。

【0041】2.  $f(u_4, c_3) = 4$

○嗜好度の予測値

【数8】

$$\hat{f}(u_4, c_3) = \frac{u(2, 4) \cdot v(3, 3) \cdot 4}{u(2, 4) \cdot v(3, 3)} = 4 \quad \dots \dots (8)$$

○重みの更新

\*【数9】

(a) (u2, c3) v. s. (u4, c3) \*

$$u(2, 4) := u(2, 4) \cdot 0.8^{(|f(u_2, c_3) - f(u_4, c_3)| - 2)v(3, 3)} = u(2, 4) \cdot 1.25^1 = 1.56.$$

. . . . . (9)

【0042】3. f(u2, c6) = 1

※【数10】

○嗜好度の予測値

※

$$\hat{f}(u_2, c_6) = \frac{u(2, 2) \cdot v(3, 6) \cdot 4 + u(2, 4) \cdot v(3, 6) \cdot 4}{u(2, 2) \cdot v(3, 6) + u(2, 4) \cdot v(3, 6)} = 4$$

. . . . . (10)

○重みの更新

★【数11】

(a) (u4, c3) v. s. (u2, c6) ★

$$u(2, 4) := u(2, 4) \cdot 0.8^{(|f(u_4, c_3) - f(u_2, c_6)| - 2)v(3, 6)} = u(2, 4) \cdot 0.8^1 = 1.25$$

$$v(3, 6) := v(3, 6) \cdot 0.8^{(|f(u_4, c_3) - f(u_2, c_6)| - 2)u(2, 4)} = v(3, 6) \cdot 0.8^1 = 0.8$$

. . . . . (11)

(b) (u2, c3) v. s. (u2, c6) ☆30☆【数12】

$$v(3, 6) := v(3, 6) \cdot 0.8^{(|f(u_2, c_3) - f(u_2, c_6)| - 2)u(2, 2)} = v(3, 6) \cdot 0.8^1 = 0.64$$

. . . . . (12)

【0043】4. f(u4, c6) = 1

◆【数13】

○嗜好度の予測値

◆

$$\begin{aligned} \hat{f}(u_4, c_6) &= \frac{u(2, 4) \cdot v(3, 6) \cdot 4 + u(4, 4) \cdot v(3, 6) \cdot 4 + u(2, 4) \cdot v(6, 6) \cdot 1}{u(2, 4) \cdot v(3, 6) + u(4, 4) \cdot v(3, 6) + u(2, 4) \cdot v(6, 6)} \\ &= \frac{1.25 \cdot 0.64 \cdot 4 + 1.0 \cdot 0.64 \cdot 4 + 1.25 \cdot 1.0 \cdot 1}{1.25 \cdot 0.64 + 1.0 \cdot 0.64 + 1.25 \cdot 1.0} \\ &= \frac{7.01}{2.69} = 2.61 \end{aligned}$$

. . . . . (13)

○重みの更新

【数14】

(a) (u2, c3) v. s. (u4, c6)



15

16

$$u(2,4) := u(2,4) \cdot 0.8^{(|f(u2,c3)-f(u4,c6)|-2)v(3,6)} = u(2,4) \cdot 0.8^{0.512} = 1.25 \cdot 0.892 = 1.115$$

$$v(3,6) := v(3,6) \cdot 0.8^{(|f(u2,c3)-f(u2,c6)|-2)u(2,4)} = v(3,6) \cdot 0.8^{1.25} = 0.512 \cdot 0.757 = 0.387$$

..... (14)

(b) (u4, c3) v. s. (u4, c6) \* \* [数15]

$$v(3,6) := v(3,6) \cdot 0.8^{(|f(u4,c3)-f(u4,c6)|-2)u(4,4)} = v(3,6) \cdot 0.8^1 = 0.512$$

..... (15)

(c) (u2, c6) v. s. (u4, c6) ※ ※ [数16]

$$u(2,4) := u(2,4) \cdot 0.8^{(|f(u2,c6)-f(u4,c6)|-2)v(3,6)} = u(2,4) \cdot 0.8^{-2.0} = 1.115 \cdot 1.56 = 1.742$$

..... (16)

【0044】以下、学習データが順次与えられる毎に重みの更新が行われ、それに応じて嗜好度の予測値の精度が高まっていく。図3に嗜好度の例（ユーザ嗜好関数の例）を示す。この例のユーザ嗜好関数では、例えば、コンテンツc1とコンテンツc3に対する各ユーザu1～u5の嗜好度が類似していることから、コンテンツc1とコンテンツc3は高い類似度であることがわかる。また、各コンテンツc1～c7に対するユーザu2の嗜好度とユーザu4の嗜好度が類似していることから、ユーザu2とユーザu4とが高い類似度であることがわかる。

【0045】次に本発明の第2の実施の形態について説明する。

【0046】上述した第1の実施の形態においては、新★

$$p = \frac{e^{u(x,x')}}{e^{u(x,x')} + e^{v(y,y')}}$$

【0048】次に本発明の第3の実施の形態について説明する。

【0049】図5を参照すると、本発明のユーザ嗜好自動学習方式の第3の実施の形態は、学習データ入力部11と、学習データ格納部12と、予測部13と、学習方式記憶部14と、学習部15と、質問部16とから構成されている。これらは、例えばプログラム制御されたCPU、キーボード等の入力装置、磁気ディスク等の記憶装置、ディスプレイ装置などを有するデータ処理装置（コンピュータ）で実現することができる。

【0050】学習データ入力部11は、学習データを逐次的に入力する部分である。入力される学習データは、ユーザ・コンテンツ対と、そのコンテンツに対するそのユーザの正しい嗜好度を表す数値とから構成される。

【0051】学習データ格納部12は過去に与えられた

★しいユーザ・コンテンツ対に対する嗜好の正解を与えられる度に、ユーザ間の重みuと、コンテンツ間の重みvとの両方を更新するが、この第2の実施の形態においては、乱数を用いてランダムにそのどちらか一方を第1の実施の形態における更新式等を用いて更新する。それ以外は第1の実施の形態と基本的に同じである。このとき、各更新を確率2分の1で行うことも可能であるが、関連する重みに依存した確率でランダムに行うことも可能である。例えば、各ユーザ間の重みu(x, x')の更新を以下の確率pで行い、ユーザ間の重みが更新されない場合（即ち確率1-p）、コンテンツ間の重みv(y, y')を更新することが考えられる。

【0047】

[数17]

..... (17)

学習データを一括して格納する部分である。

【0052】学習方式記憶部14は、複数の学習方式を格納する部分である。ここで、各学習方式は、与えられたユーザ・コンテンツ対に対して、そのユーザのそのコンテンツに対する嗜好度を現在の仮説に基づいて予測する機能と、そのユーザのそのコンテンツ対に対する正しい嗜好度に基づき現仮説を更新（学習）する機能とを持つ。

【0053】予測部13は、学習データ入力部11から入力されたユーザ・コンテンツ対について、そのユーザのそのコンテンツに対する嗜好度を予測する部分である。この予測は、そのユーザのそのコンテンツに対する嗜好度を、学習方式記憶部14に格納された各学習方式により予測させ、各学習方式に対する現在の信頼度を表す重みを用いて、それらの予測値の重み付き平均値をと

ることで行われる。

【0054】質問部16は、システム自らユーザ・コンテンツ対を指定した質問を出し、そのユーザ・コンテンツ対に対する嗜好度の入力を受け取る部分である。質問は例えばディスプレイの画面に出され、回答はキーボードや画面上から入力される。

【0055】学習部15は、学習方式記憶部14に記憶された各学習方式の現仮説および各学習方式の信頼度を表す重みを更新する部分である。

【0056】図6は本実施の形態の動作の一例を示すフローチャートである。この動作例では、学習データ入力部11から1つのユーザ・コンテンツ対を入力し、そのコンテンツに対するそのユーザの嗜好度を予測部13において予測し、次いで学習データ入力部11からそのユーザ・コンテンツ対の正しい嗜好度を入力して学習部15において各学習方式の現仮説および各学習方式の信頼度を表す重みを更新し、そして、質問部16において可能であれば1つのユーザ・コンテンツ対の嗜好度を質問し、その回答に応じて学習部15において各学習方式の現仮説および各学習方式の信頼度を表す重みを更新する、という処理を繰り返す逐次式の学習を行う。以下、本実施の形態について詳述する。

【0057】いま、ユーザの集合をX、コンテンツの集合をYとすると、本実施の形態においては、図6に示されるように、以下のようなステップを実行する。

(1) 新しいユーザ $x \in X$ と新しいコンテンツ $y \in Y$ の対を学習データ入力部11にて入力する(ステップS11)。

(2) 入力されたユーザ $x$ とコンテンツ $y$ の対について、ユーザ $x$ のコンテンツ $y$ に対する嗜好度 $f(x, y)$ を、予測部13が学習方式記憶部14に格納された複数の学習方式による予測値および各学習方式の現在の信頼度を表す重みを用いて予測する(ステップ12, 13)。 $f(x, y)$ の値は一般に実数値であると仮定する。例えば、5段階評価(1, 2, ..., 5)等が想定される。

(3) ユーザ $x$ のコンテンツ $y$ に対する正しい嗜好度 $f(x, y)$ の値を学習データ入力部11にて入力し(ス\*

$$\hat{f}(x, y) = \frac{\sum_i w_i A_i(x, y)}{\sum_i w_i}$$

【0061】なお、 $(x, y)$ に対する正しい嗜好度 $f(x, y)$ が入力として与えられたとき、各学習方式 $A_i$ の重み $w_i$ は以下のように更新される。

$$w_i := w_i \cdot e^{c \cdot |f(x, y) - A_i(x, y)|}$$

【0063】以上のような重み更新により、学習が進むと共に、より信頼度の高い学習方式の予測値が大きい影響力を持つようになる。

【0064】次に、少ない数の質問数での高い精度の予

\* テップS14)、学習データ格納部12に格納する(ステップS15)。

(4) ステップS14で入力された正しい嗜好度に基づき、学習部6において複数の学習方式の現在の仮説を更新(学習)し(ステップS16)、また、各学習方式の信頼度を表す重みを更新する(ステップS17)。

(5) 質問部16において、複数の学習方式の現在の仮説を用いて、嗜好度を質問すべき新たなユーザ $x'$ とコンテンツ $y'$ の対を求め、その対に対する正しい嗜好度をユーザから入力として得る(ステップS18)。

(6) 上記ユーザ $x'$ とコンテンツ $y'$ の対、その正しい嗜好度から構成される学習データを学習データ格納部12に格納する(ステップS19)。

(7) ステップS18で入力された正しい嗜好度に基づき、学習部6において複数の学習方式の現在の仮説を更新(学習)し(ステップS20)、また、各学習方式の信頼度を表す重みを更新する(ステップS21)。

(8) 学習を終了するか否かを判定し、終了しない場合はステップS11に戻る(ステップS22)。

【0058】予測・学習アルゴリズムの目標は、なるべく少ない数の学習データおよび質問数で、高い精度の予測を行うことにある。ステップS22で終了とする条件(停止条件)としては、或る一定の学習精度が達成されたこと等の様々な条件が考えられるが、本発明については本質的な影響はなく、また一般に適用環境に大きく依存するので、ここでは特に特定しない。なお、上記

(1)～(7)を際限なく繰り返すようにしても良い。

【0059】本実施の形態において、少ない学習データによる高い精度の予測に関しては、学習方式記憶部14に記憶された複数の学習方式にそれぞれ与えられたユーザ・コンテンツ対における嗜好度を予測させて、それらの学習方式の信頼度を表す重みを用いた重み付け平均等で予測を行うことで達成される。すなわち、各学習方式による予測値を $A_i(x, y)$ 、その重みを $w_i$ として、 $f(x, y)$ の値を以下の $\hat{f}(x, y)$ をもって予測する。

【0060】

【数18】

$$\dots\dots\dots (18)$$

※【0062】

【数19】

$$\dots\dots\dots (19)$$

測を行うためには、情報量の高いユーザ・コンテンツ対についてその嗜好度を質問することが必要になる。このために、学習方式記憶部14に記憶された複数の学習方式の予測値がばらついているユーザ・コンテンツ対を発

見し、その対に対する嗜好度を質問する。或るユーザ・コンテンツ対に対する予測値のばらつきは、予測値の分散または予測値の重み付き平均嗜好度に対する自乗誤差の、各学習方式の重みを用いた重み付き平均等によって測ることができる。予測値のばらつきの多い対を発見する方法については、例えば乱数を用いて発生された多数のユーザ・コンテンツ対の中から、予測値が上記の意味\*

$$(x^*, y^*) := \arg \max_{(x, y) \in P} \sum_i w_i (A_i(x, y) - \sum_i w_i A_i(x, y))^2 \quad \dots \dots (20)$$

【0066】なお、求めたユーザ・コンテンツ対が既に学習データに存在する場合には、次点以降の候補の中から学習データに存在しない対の候補の質問を選択すれば良い。

【0067】次に第4の実施の形態について説明する。

【0068】第4の実施の形態は、第3の実施の形態において、学習方式記憶部14に記憶する複数の学習方式として、第2の実施の形態のユーザ嗜好自動学習方式を実現するアルゴリズムの多数のコピーを用いる。すなわち、それぞれ同一の学習方式を用いているが、一般に、ユーザ間重み、コンテンツ間重みの更新にあたって異なる乱数が発生され、異なる重みの更新が行われていくので、特定の場面において異なる予測を行う複数の学習方式となる。特に、乱数が発生されるのは、 $(x, y)$ と $(x', y')$ が類似していないという命題から、 $(x, x')$ が類似していないという命題と $(y, y')$ が類似していないという命題のどちらを結論づけるかの曖昧性を解消するために行われるので、これらの曖昧性に関して異なった結論を下した多数のコピーが共存することになる。それらの予測値の、各学習方式の信頼度を表す重みによる重み付き平均をもって実際の嗜好度の予測を行うことにより、より高い精度の学習が可能になるし、またそれらの予測値のばらつきを最大にするユーザ・コンテンツ対に対して質問を行うことにより情報量の多い質問を実現することができる。

【0069】次に第5の実施の形態について説明する。

【0070】第5の実施の形態においては、第4の実施の形態と同様に、学習方式記憶部14に記憶する複数の学習方式として、第2の実施の形態のユーザ嗜好自動学習方式を実現するアルゴリズムの多数のコピーを用いる。また、第3および第4の実施の形態においては、予測値のばらつきの多い対の発見については、多数の対を生成し比較する方法を採用したが、この第5の実施の形態においては、この予測値のばらつきの多い対の発見について或る特定の方法を用いる。

【0071】すなわち、複数のコンテンツとユーザの対からなる質問候補集合として、過去の学習データ中に現れるユーザとコンテンツからなるユーザ・コンテンツ対の中で、過去にデータとして嗜好度を与えられていない対の集合の全て又は一部を用いて、それらの中で最も予測値のばらつきの大きいものについて質問を行う方法

\*で最もばらついている対を選んで質問するといった方法が採用できる。すなわち、質問に用いるユーザ・コンテンツ対 $(x', y')$ は、候補として考慮されるユーザ・コンテンツ対の集合をPとして、以下のように求めることができる。

【0065】

【数20】

である。

【0072】この方法は、 $f(x, y)$ と $f(x', y')$ が与えられて、その差が大きいことが判明したとき生じる曖昧性、すなわち $x, x'$ の類似度が低いのか、 $y, y'$ の類似度が低いのかの曖昧性の具体的な解消法として、 $f(x, y')$ または $f(x', y)$ の値を質問により知ることが有効であるという知見に基づいている。すなわち、そのような対に対する質問は情報量が多いので、その中で最も予測値のばらつきの大きい対を選択して質問することにより、確実に情報量の大きい質問を実現することが可能になる。

【0073】図7に第5の実施の形態の構成例を示す。同図に示すように、本実施の形態は、学習データを逐次的に入力する学習データ入力部11と、過去に与えられた学習データを一括して格納する学習データ格納部12と、複数の学習方式14-1, 14-2, ...を格納する学習方式記憶部14と、学習データ入力部11から入力されたユーザ・コンテンツ対について、その嗜好度を各学習方式14-1, 14-2, ...に予測させ、各学習方式の信頼度を表す重みを用いて、それらの予測値の重み付き平均値をとり、それを当該ユーザ・コンテンツ対の予測値とする予測部13と、システム自らユーザ・コンテンツ対を指定した質問を出し、そのユーザ・コンテンツ対に対する嗜好度の入力を受け取る質問部16と、学習方式記憶部14に記憶された各学習方式14-1, 14-2, ...の現仮説および各学習方式の信頼度を表す重みを更新する学習部15とから構成されている。

【0074】そして、各学習方式14-1, 14-2は、図1で説明したようなユーザ間重み記憶部3、コンテンツ間重み記憶部4、予測部5および学習部6で構成されている。第2の実施の形態のユーザ嗜好自動学習方式の実装なので、学習部6は、乱数を用いてランダムに、ユーザ間重み又はコンテンツ間重みを更新する。

【0075】なお、この第5の実施の形態の動作例は図6に示したフローチャートと同じである。

【0076】図7において、利用者から1つのユーザ・コンテンツ対が学習データ入力部11に入力されると(ステップS11)、学習データ入力部11は、それを予測部13に伝達する。予測部13は、伝達されたユーザ・コンテンツ対を各学習方式14-1, 14-2, ...に与え、その嗜好度を予測させる。

【0077】各学習方式14-1, 14-2, …においては、その各々の予測部5が、学習データ格納部12に格納された学習データ、自方式内のユーザ間重み記憶部3に格納されたユーザ間重み及びコンテンツ間重み記憶部4に格納されたコンテンツ間重みを用いて、そのユーザのそのコンテンツに対する嗜好度を予測する(ステップS12)。つまり、式(1)に基づき、学習データ格納部12に格納される過去の学習データの嗜好度の値の、新たに与えられたユーザ・コンテンツ対との間のユーザ間重みとコンテンツ間重みの積を重みとする、重みつき平均を予測値とし、予測部13に返却する。

【0078】予測部13は、各学習方式14-1, 14-2, …の予測値と、各学習方式14-1, 14-2, …の現在の信頼度を表す重みとを用いて、前述した式18によって予測値を計算する(ステップS13)。

【0079】次いで、学習データ入力部11から上記ユーザ・コンテンツ対の正しい嗜好度が入力されると(ステップS14)、それが学習データ格納部12に格納されると共に(ステップS15)、学習部15に伝達される。学習部15は、伝達された嗜好度を各学習方式14-1, 14-2, …の学習部6に伝達し、各々のユーザ間重み、コンテンツ間重みを更新させる(ステップS16)。このとき、各学習方式14-1, 14-2, …における学習部6は、前述した式(17)の確率で前述した式(2)等によってユーザ間重みを更新し、ユーザ間重みを更新しない場合は前述した式(3)等によってコンテンツ間重みを更新する。

【0080】学習部15は、各学習方式14-1, 14-2, …における学習と同時に、予測部13が使用する各学習方式の信頼度を表す重みを前述した式(19)によって更新する(ステップS17)。

【0081】次に、質問部16は、複数のユーザ・コンテンツ対からなる質問対候補集合として、学習データ格納部12中に現れるユーザとコンテンツとからなるユーザ・コンテンツ対の中で、過去にデータとして嗜好度を与えられていない対の集合の全て又は一部を用い、その集合中の各対について各学習方式14-1, 14-2, …にその嗜好度を予測させ、その予測値のばらつきの最も大きいユーザ・コンテンツ対の嗜好度を利用者に質問する。なお、このとき各学習方式14-1, 14-2, …は予測部13からユーザ・コンテンツ対が与えられた場合と同様にして予測を行う。次いで、質問部16は、\*

$$v(3,6) := v(3,6) \cdot 0.8^{(|f(u4,c3) - f(u2,c6)| - 2)u(2,4)} = v(3,6) \cdot 0.8^1 = 0.8$$

..... (21)

【0085】○学習方式14-2

1.  $f(u2, c3) = 4$

重みの更新なし。

2.  $f(u4, c6) = 1$

【数22】

\* 質問したユーザ・コンテンツ対に対して利用者から正しい嗜好度が入力されると、それを学習データ格納部12に格納すると共に(ステップS19)、それを学習部15に伝達する。

【0082】学習部15は、伝達された学習データの嗜好度を各学習方式14-1, 14-2, …の学習部6に伝達し、各々のユーザ間重み、コンテンツ間重みを更新させる(ステップS20)。このとき、各学習方式14-1, 14-2, …における学習部6は、前述した式(17)の確率で前述した式(2)等によってユーザ間重みを更新し、ユーザ間重みを更新しない場合は前述した式(3)等によってコンテンツ間重みを更新する。同時に学習部15は、予測部13が使用する各学習方式の信頼度を表す重みを前述した式(19)によって更新する(ステップS21)。

【0083】次に、嗜好度の予測および重み行列の更新の具体例を説明する。例としては、簡単のために、学習方式記憶部14に記憶された学習方式は、14-1と14-2の2つとする。また、図4に示すようにu1からu5までの5人のユーザとc1からc7までの7つのコンテンツとを想定し、学習データ格納部12における過去の学習データを示すデータ行列Mの要素は全て未設定(図4では?で示す)とする。更に、各学習方式14-1, 14-2のユーザ間重み記憶部3に設けられているユーザ間重み及びコンテンツ間重み記憶部4に設けられているコンテンツ間重みはそれぞれ初期値1に設定されているものとする。更に、定数cは、 $-1 \log 0.8$ に設定、すなわち $e^{-c} = 0.8$ であり、dは2であるとする。また、自分自身の重み $u(x, x)$ 、 $v(y, y)$ は更新しないものとする。このとき、学習データとして、 $f(u2, c3) = 4$ が与えられ、この時点では質問する対の候補が空集合なので質問はされず、次に学習データ $f(u4, c6) = 1$ が与えられたとする。これらのデータに対して、学習方式14-1, 14-2におけるユーザ間重み又はコンテンツ間重みは、例えば以下のようになる。

【0084】○学習方式14-1

1.  $f(u2, c3) = 4$

重みの更新なし。

2.  $f(u4, c6) = 1$

【数21】

$$u(2,4) := u(2,4) \cdot 0.8(|f(u4,c3) - f(u2,c6)| - 2)v(3,6) = u(2,4) \cdot 0.8^1 = 0.8$$

..... (22)

【0086】つまり、 $f(u4, c6) = 1$ の学習データに対して、学習方式14-1ではコンテンツ間重みを更新し、学習方式14-2ではユーザ間重みを更新している。

\* する嗜好度の学習方式14-1、14-2による予測値は以下のようになる。

1.  $f(u2, c6)$   
(a) 学習方式14-1

【0087】この段階において、質問する対の候補は、10 【数23】  
(u2, c6)と(u4, c3)であり、その各々に対\*

$$\begin{aligned} \hat{f}(u2, c6) &= \frac{u(2,2) \cdot v(3,6) \cdot 4 + u(2,4) \cdot v(6,6) \cdot 1}{u(2,2) \cdot v(3,6) + u(2,4) \cdot v(6,6)} \\ &= \frac{1.0 \cdot 0.8 \cdot 4 + 1.0 \cdot 1.0 \cdot 1}{1.0 \cdot 0.8 + 1.0 \cdot 1.0} = \frac{4.2}{1.8} = 2.33 \quad \dots \dots (23) \end{aligned}$$

(b) 学習方式14-2

※ ※ 【数24】

$$\begin{aligned} \hat{f}(u2, c6) &= \frac{u(2,2) \cdot v(3,6) \cdot 4 + u(2,4) \cdot v(6,6) \cdot 1}{u(2,2) \cdot v(3,6) + u(2,4) \cdot v(6,6)} \\ &= \frac{1.0 \cdot 1.0 \cdot 4 + 0.8 \cdot 1.0 \cdot 1}{1.0 \cdot 0.8 + 1.0 \cdot 1.0} = \frac{4.8}{1.8} = 2.67 \quad \dots \dots (24) \end{aligned}$$

【0088】2.  $f(u4, c3)$

★ 【数25】

(a) 学習方式14-1

★

$$\begin{aligned} \hat{f}(u4, c3) &= \frac{-u(2,4) \cdot v(3,3) \cdot 4 + u(4,4) \cdot v(3,6) \cdot 1}{u(2,4) \cdot v(3,3) + u(4,4) \cdot v(3,6)} \\ &= \frac{1.0 \cdot 1.0 \cdot 4 + 1.0 \cdot 0.8 \cdot 1}{1.0 \cdot 1.0 + 1.0 \cdot 0.8} = \frac{4.8}{1.8} = 2.67 \quad \dots \dots (25) \end{aligned}$$

(b) 学習方式14-2

☆ ☆ 【数26】

$$\begin{aligned} \hat{f}(u4, c3) &= \frac{u(2,4) \cdot v(3,3) \cdot 4 + u(4,4) \cdot v(3,6) \cdot 1}{u(2,4) \cdot v(3,3) + u(4,4) \cdot v(3,6)} \\ &= \frac{0.8 \cdot 1.0 \cdot 4 + 1.0 \cdot 1.0 \cdot 1}{0.8 \cdot 1.0 + 1.0 \cdot 1.0} = \frac{4.2}{1.8} = 2.33 \quad \dots \dots (26) \end{aligned}$$

【0089】この場合、両候補対に対する両学習方式14-1、14-2の予測値は、それぞれ2.67, 2.33と、2.33, 2.67であるので、両対とも同等にばらつきを示しており、この段階においては、両対とも同等に情報量が多いので、任意にどちらかの対を選択し質問を行う。ここでは、(u2, c6)が選択されて質問が行われたと仮定する。

◆ 1, 14-2による上記のデータに対する予測値と、その結果行われる各学習方式の信頼度を表す重みの更新を以下に示す。なお、以下の例において、各学習方式 $A_i$ の重み $w_i$ の初期値は1.0とし、以下のように更新する。また、最初に与えられる対に対する予測は、嗜好度の中間点である3点とする。

40 【0091】

【0090】次に、この段階までの各学習方式14-

◆ 【数27】

$$w_i := w_i \cdot e^{\log 0.8 \cdot |f(x,y) - A_i(x,y)|} = w_i \cdot 0.8^{|f(x,y) - A_i(x,y)|} \quad \dots \dots (27)$$

【0092】○学習方式14-1において

\* 【数28】

1.  $A_1(u2, c3) = 3$

\*

$$w_1 := w_1 \cdot 0.8^{|f(u2,c3) - A_1(u2,c3)|} = 1.0 \cdot 0.8^{4-3} = 0.8 \quad \dots \dots (28)$$

2.  $A_1(u4, c6) = 4$

50 【数29】

25

26

$$w_1 := w_1 \cdot 0.8^{|f(u4,c6) - A_1(u4,c6)|} = 0.8 \cdot 0.8^{|1-4|} = 0.410 \quad \dots \dots (29)$$

3.  $A_1(u2, c6) = 2.33$  \* \* 【数30】

$$w_1 := w_1 \cdot 0.8^{|f(u2,c6) - A_1(u2,c6)|} = 0.410 \cdot 0.8^{|1-2.33|} = 0.305$$

\dots \dots (30)

【0093】○学習方式14-2において ※【数31】

1.  $A_1(u2, c3) = 3$  ※10

$$w_2 := w_2 \cdot 0.8^{|f(u2,c3) - A_1(u2,c3)|} = 1.0 \cdot 0.8^{|4-3|} = 0.8 \quad \dots \dots (31)$$

2.  $A_1(u4, c6) = 4$  ★ ★ 【数32】

$$w_2 := w_2 \cdot 0.8^{|f(u4,c6) - A_1(u4,c6)|} = 0.8 \cdot 0.8^{|1-4|} = 0.410$$

\dots \dots (32)

3.  $A_1(u2, c6) = 2.33$  ☆ ☆ 【数33】

$$w_2 := w_2 \cdot 0.8^{|f(u2,c6) - A_1(u2,c6)|} = 0.410 \cdot 0.8^{|1-2.33|} = 0.282$$

\dots \dots (33)

【0094】上記の例において、 $f(u2, c3) = 4$  および  $f(u4, c6) = 1$  が学習データとして与えられた段階で、 $u2$ 、 $u4$  または  $c3$ 、 $c6$  のいずれかの類似度が低い、どちらかはわからないという曖昧性が存在している。この曖昧性を、学習方式14-1と学習方式14-2においてランダムに解消しているが、上記の例においては、学習方式14-1においては、 $v(3, 6)$  を減らす方向に更新しており、学習方式14-2においては、 $u(2, 4)$  を減らす方向に更新している(図3によれば $u2$ 、 $u4$  は高い類似度をもっている)。この結果、上記の例において選択された質問点に対する両学習方式による予測値の誤差に差が生じ、結果的に正しい曖昧性解消を行った学習方式14-1の方が、重みが大きくなっており(すなわち  $w_1 > w_2$ )、次に与えられるユーザ・コンテンツ対に対する予測においては、学習方式14-1の予測値の方が重んじられることが分かる。

【0095】

【発明の効果】請求項1または2記載のユーザ嗜好自動学習方式によれば、ノイズに対して頑健な帰納推論を実現することが可能になると同時に、ユーザ間の類似度のみを考慮する場合や、コンテンツ間の類似度のみを考慮

する従来方法に比べて、立上がりの早い学習が可能になる。

30 【0096】また、請求項3、4または5記載のユーザ嗜好自動学習方式によれば、従来方法と比べて格段に少ない数の質問数で、精度の良い学習を達成することが可能になる。

【図面の簡単な説明】

【図1】本発明のユーザ嗜好自動学習方式の第1の実施の形態の構成例を示すブロック図である。

【図2】本発明の第1の実施の形態の動作の一例を示すフローチャートである。

【図3】嗜好度(ユーザ嗜好関数)の例を示す図である。

【図4】過去の学習データの例を示す図である。

【図5】本発明のユーザ嗜好自動学習方式の第3の実施の形態の構成例を示すブロック図である。

【図6】本発明の第3の実施の形態の動作の一例を示すフローチャートである。

【図7】本発明の第5の実施の形態の構成例を示すブロック図である。

【符号の説明】

1…学習データ入力部

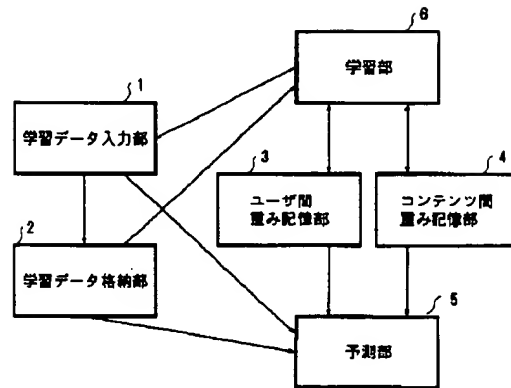
2…学習データ格納部

- 3...ユーザ間重み記憶部  
 4...コンテンツ間重み記憶部  
 5...予測部  
 6...学習部  
 11...学習データ入力部  
 12...学習データ格納部

- \* 13...予測部  
 14...学習方式記憶部  
 14-1, 14-2...学習方式  
 15...学習部  
 16...質問部

\*

【図1】



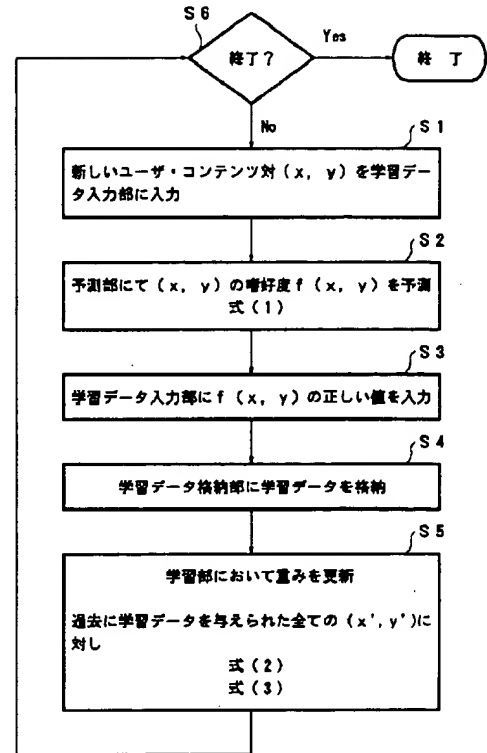
【図3】

高い類似度

	c1	c2	c3	c4	c5	c6	c7
u1	1	4	1	4	5	2	4
u2	5	2	4	3	3	1	2
u3	1	1	1	2	4	5	2
u4	4	2	4	3	3	1	3
u5	5	4	5	3	2	1	3

高い類似度

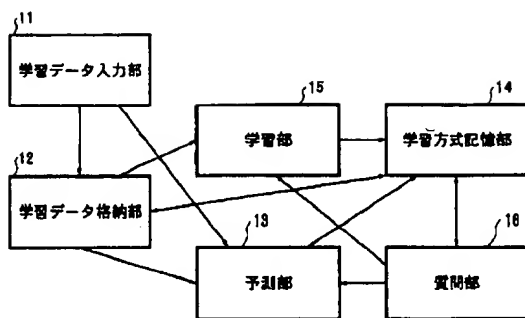
【図2】



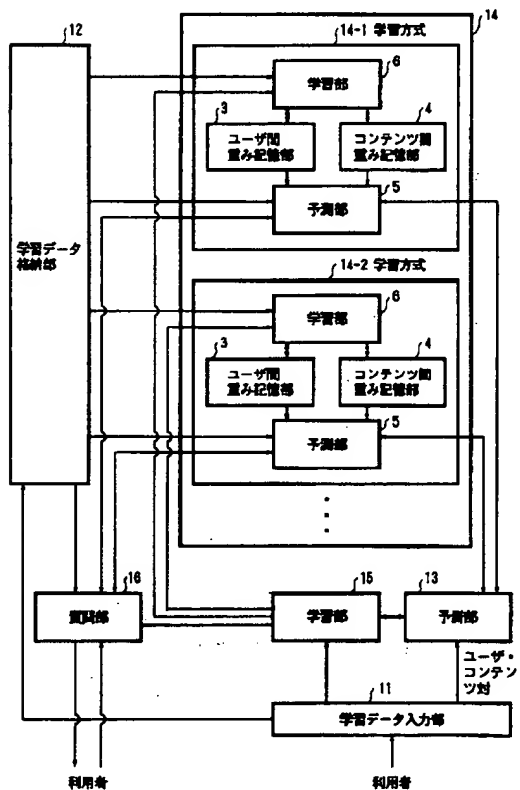
【図4】

	c1	c2	c3	c4	c5	c6	c7
u1	?	?	?	?	?	?	?
u2	?	?	4	?	?	1	?
u3	?	?	?	?	?	?	?
u4	?	?	4	?	?	①	?
u5	?	?	?	?	?	?	?

【図5】



【図7】



【図6】

